University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

School of Information Sciences -- Faculty Publications and Other Works

School of Information Sciences

7-1-1995

# Picking the Best Databases

Carol Tenopir
*University of Tennessee - Knoxville*

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs

 Part of the Collection Development and Management Commons

## Recommended Citation

Tenopir, Carol, "Picking the Best Databases" (1995). *School of Information Sciences -- Faculty Publications and Other Works.*
https://trace.tennessee.edu/utk_infosciepubs/378

This Article is brought to you for free and open access by the School of Information Sciences at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

# ONLINE DATABASES

BY CAROL TENOPIR

# Picking the Best Databases

WHEN I STARTED ONLINE searching 20 years ago I knew almost every database intimately. With only 300 databases, it wasn't too difficult to have confidence in my ability to remember which ones were the best. ERIC for education or training questions, BIOSIS for biological queries, MEDLINE for medicine—the choices were straightforward and clearcut. I knew which ones were updated on a timely basis, which kept their indexing up-to-date, and which covered the important literature in a field.

According to the *Gale Directory of Databases*, there are now almost 9000 commercially available databases on about 1700 systems. Add to that the tens of thousands of sources available from thousands of Internet sites, and the task of selecting the best seems nearly impossible.

## Automatic finding tools

Sophisticated finding tools such as DIALINDEX on DIALOG or the Internet-based Webcrawler or Lycos narrow down database choices by calculating how often the words in your query appear in each database. For a given question they allow you to identify likely sources by identifying those that have the most information on your topic, or at least have the most occurrence of your search words.

But such aids work only for a given query and do not address the more subtle characteristics of database quality. All of the evaluation criteria that librarians learn in collection development are more subtle than word counting; scope, uniqueness, currency, audience, style, accuracy, arrangement, authority, and so forth. None of the automatic tools tell you whether a database is the best or most important in a general subject area.

Carol Tenopir is Professor at the School of Library and Information Science, University of Tennessee at Knoxville. Her E-mail address is tenopir@utkux.utk.edu

## Subjective process

The subtleties of database evaluation still require a human brain. Like most things analyzed by humans, the process is subjective even when consistent criteria are applied. It relies on extensive experience searching many databases—a frustrating fact for beginning searchers.

Experienced searcher Mick O'Leary, library director at Frederick Community College in Maryland, has written an evaluative guide to the best databases. *The Online 100: ONLINE Magazine's Field Guide to the 100 Most Important Online Databases* dares to select the best databases in broad subject categories. He has verified his choices through an expert advisory group of eight other experienced searchers. These are the 100 most important commmercially available online databases chosen by consensus of a powerful group of online experts.

## Database choices

Database choices are divided into 10 subject categories: 1. General & Business News; 2. Business; 3. U.S. and International Companies; 4. Current Events, Law and Government; 5. General and Physical Science; 6. Life Sciences & Medicine; 7. Technology & Computing; 8. Intellectual Property; 9. Social Sciences & Humanities; and 10. General Reference/Multidisciplinary. Each section includes between six and 14 database choices, for a total of 100 first choices. There are actually more than 100 total databases listed, because each section includes two, three, or four "honorable mention" databases.

I was pleased to see that many of my favorite databases (selected from experience and gut instinct) were also selected by O'Leary and his advisors. Many of his choices are the old standbys whose quality has endured for many years, including: ABI/INFORM and PROMPT for business; DISCLOSURE SEC and Dun's Market Identifiers for U.S. and International Companies; Magazine Database, PAIS International, LEXIS, and WESTLAW for Current Events, Law, and Government; SciSearch, CA File, and INSPEC for General and Physical Science; BIOSIS Previews, CAB Abstracts, AGRICOLA, and MEDLINE for Life Sciences & Medicine; Compendex Plus and NTIS for Technology & Computing; SciSearch, PsycINFO, ERIC, and Historical Abstracts for Social Sciences & Humanities; and Books in Print and Dissertation Abstracts for General Reference/Multidisciplinary.

Other choices were more surprising. I will be sure to take another look at the recommended "A Matter of Fact" and "ArticleFirst/ContentsFirst" on EPIC and FirstSearch; "Political Risk Services," on DataStar, LEXIS-NEXIS, and NewsNet; and "Knight-Ridder/Tribune Business News" on DIALOG.

Defining what a database is can be difficult enough, let alone choosing the best ones. Thus LEXIS, NEXIS, NewsNet and WESTLAW are each counted as one database, when in reality they are complete online systems. DIALOG PAPERS gets counted as a database, when it is really DIALOG's OneSearch search feature that allows the user to search dozens of independent newspaper files simultaneously. Dow Jones Text Library is also counted as one database.

We would all agree that separate files such as Magazine Database or Newspaper & Periodical Abstracts are true databases—they made the list competing with the likes of LEXIS and WESTLAW. Among separate files listed, Information Access Company wins the honors as the most prolific producer of important databases, with ten of the top 100. Dun & Bradstreet produces five of the selected titles, UMI and Knight-Ridder produce four each, and both H.W. Wilson and Institute for Scientific Information produce three.

## Information about each database

For each of the top 100 databases, O'Leary provides about one and a half

pages of concise information. He begins with a brief introduction that puts each database into context.

Each database is summarized in consistently labeled sections. "Content Notes," the longest section, typically 150-200 words, explains what each database covers, lists which topics it is strong in, and compares it to rival databases. "Content Notes" includes the kind of information typically found in an online system's database summary sheets or in introductions to database search manuals, with the added value of how this database stands apart from similar ones.

"Search Notes" are system independent hints for using each database better. In the Reuters entry, for example, O'Leary points out that since the stories are not indexed, full text search techniques are important. For INSPEC he recommends becoming familiar with the "extremely thorough indexing" found in descriptors, identifiers, classification names and codes, document type and article treatment, and specialized chemical and numerical indexing. He reminds searchers that INSPEC uses British spelling.

Since many search features vary with the system on which a database is loaded, search hints cannot be terribly specific. O'Leary tries to provide search hints that will be true across all online systems. The entries that do mention specific system features, however, most often refer to DIALOG. Entries for Magazine Database, Computer Database, and Trade & Industry Database, for example, describe how DIALOG places daily updates for all of these databases into the merged file Newsearch. Entries for some directory databases recommend DIALOG's REPORT command.

## "Do not use for..."

An especially useful section is the "Do Not Use For..." hints. Equally important are the best applications for the database, information on the limitations of a database, or pointers to other databases in the guide that may be better suited for a particular purpose. INSPEC, for example, should not be used for "lay-oriented treatment of computing subjects and applications." O'Leary recommends Computer Databases or Magazine Databases instead. Reuters should not be used for "most U.S. news searching... Among comprehensive newswires, the Associated Press has more U.S. coverage, especially for local news."

Some entries include brief tidbits under "Did You Know That..." This additional information varies, from details about the company that produces the database, to more explanations about the content, or simply interesting historical facts. The Reuters entry, for example, reads "Did You Know That... Reuters invented the wire service? Its founder, Paul Julius Reuter, began transmitting stock quotations between London and Paris in 1849."

## Guide for key facts

Each database entry ends with a "Key Facts" section, which includes seven facts, each presented in just one or two lines. Some of these are more complete than others. First is a categorization of "type" (bibliographic, abstract, full text, directory, numeric, or a combination of these). Next is "span," a brief mention of each database's date coverage. Since date coverage may vary from system to system and sometimes even within one system, date span cannot be entirely accurate. The span for Magazine Database, for example, is given just as "1959 to date." The entry doesn't mention Magazine Index's two-and-a-half-year gap in indexing in the early 1970s, nor does it explain that the full text records began only with 1983. Elsewhere in the Magazine Database entry, it does explain that abstracts were included only since 1992.

The database producer's name, address, and phone number are given for each entry, as are the "Principle Databanks" (online systems) that carry the database, the 22 principal online systems' addresses which are repeated in Appendix B and include major consumer online systems such as America Online, CompuServe, Prodigy, and Delphi, as well as systems used more by information professionals, such as DIALOG, OCLC, EPIC, LEXIS-NEXIS, STN, etc.

The next category of key fact, "Available on the Internet," may be more useful in later editions of the book. Only a few of the top 100 entries contain anything except "no" to this line. Of those that do contain an entry, many point to documentation or database producer web homepages. Only FedWorld, Cendata, and Britannica Online have true Internet versions (Britannica is only available on Internet and requires a subscription fee). The entries also provide Internet addresses. Entries for DIALOG's U.S. Copyrights and U.S. Patents Full-text databases only tell searchers that copyrights and patents are available on Internet from the Copyright or U.S. Patent and Trademark offices, leaving them to locate the addresses themselves.

The information provided under the CD-ROM versions of databases is even less consistent and of limited use. In a few instances, the header doesn't even appear in entries; other times the information given is simply "no." When information is given it is usually only the name of the CD-ROM vendor(s) and even this information isn't entirely reliable or complete. *The Online 100* should not be used as a guide to CD-ROMs!

Each database entry ends with information on "Typical Search Cost." Of course this is risky for databases that are available on many different online systems, and O'Leary cautions in his introduction that, "The figure given here translates and merges pricing information across all host services and arrives at an estimated dollar cost for a very common kind of search in that database. It is calculated for a ten-minute search for a representative number of records, and is meant to be no more than an approximation." His approximations seem realistic and provide valuable warnings for experienced searchers as well as novices.

When there is a significant difference in cost between systems such as when a database is available on both research systems and a consumer system such as Knowledge Index, a "best price" approximation is given in addition to the typical cost.

## Pemberton press

*The Online 100* is useful for new intermediaries, experienced searchers, and end user researchers and, at $22.95 per copy, should have a relatively wide audience. It is the first in a planned series of moderately priced search aids by Online Inc.'s Pemberton Press. Coming in October will be *Naked in Cyberspace: How to Find Personal Information Online* by Carole A. Lane (billed as *Online Magazine*'s field guide to what you can find out online about anyone). Both can be used in conjunction with Reva Basch's *Secrets of the Super Searchers* published by Online, Inc. in 1993.

For more information contact Pemberton Press, Online Inc., 462 Danbury Road, Wilton, CT 06897-2126, (800) 248-8466 or (203) 761-1466; fax (203) 761-1444.

Copyright of Library Journal is the property of Library Journals, LLC and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.